



Received: 01/02/2025

Accepted: 28/02/2025

Anales de Edificación

Vol. 11, Nº1, 58-61 (2025)

ISSN: 2444-1309

DOI: 10.20868/ade.2025.5516

## Funciones de probabilidad de distribución para modelos aplicados a obras hidráulicas en zonas urbanas: LOG-GUMBEL vs LOG-PEARSON III

### Distribution probability functions for models applied to hydraulic works in urban areas: LOG-GUMBEL vs LOG-PEARSON III

Raúl Montes Pajuelo<sup>a</sup>; Ángel Mariano Rodríguez Pérez<sup>b</sup>; Julio José Caparros Mancera<sup>a</sup>; Cesar Antonio Rodríguez González<sup>a</sup>

<sup>a</sup>University of Huelva, Spain. raul.montes@dci.uhu.es

<sup>b</sup>University of Almería, Spain

**Resumen**-- Los modelos hidrometeorológicos para el diseño hidráulico se basan en funciones de probabilidad para modelizar las precipitaciones extremas. Este estudio compara las funciones Log-Gumbel y Log-Pearson III utilizando datos del embalse de Almodóvar (Cádiz). Log-Gumbel, una variación de la distribución de Gumbel, produce estimaciones más altas y seguras para periodos de retorno largos (T), pero es muy sensible a los cambios en los datos. Log-Pearson III, ampliamente utilizada y más estable, muestra resultados consistentes a pesar de la variabilidad de los datos. Para diseños con datos limitados o variables -comunes en obras hidráulicas- se prefiere Log-Pearson III. Sin embargo, para el drenaje urbano en el que la seguridad tiene prioridad sobre la estabilidad, Log-Gumbel ofrece una alternativa conservadora para valores de T elevados.

**Palabras clave**— Modelling; hydrological engineering; precipitation distribution functions; drainage in urban areas.

**Abstract**— Hydrometeorological models for hydraulic design rely on probability functions to model extreme rainfall. This study compares Log-Gumbel and Log-Pearson III functions using data from Almodóvar reservoir (Cádiz). Log-Gumbel, a variation of the Gumbel distribution, yields higher, safer estimates for long return periods (T), but is highly sensitive to data changes. Log-Pearson III, widely used and more stable, shows consistent results despite data variability. For designs with limited or variable data—common in hydraulic works—Log-Pearson III is preferred. However, for urban drainage where safety is prioritized over stability, Log-Gumbel offers a conservative alternative for high T values.

**Index Terms**— Modelización; ingeniería hidrológica; funciones de distribución de las precipitaciones; drenaje en zonas urbanas.

#### I. INTRODUCTION

EXTREME weather events are an issue of current interest to society due to their impact on the population and the economy. In particular, the focus is on extreme precipitation events and, more specifically, their modelling (Coronado-Hernández et al, 2020; Hossain et al, 2010; Teng et al, 2018).

This work presents a comparative analysis between the Log-Gumbel (Gumbel, 1935; Gumbel, 1941; Heo and Salas, 1996) and Log-Pearson III (Bobee and Robitaille, 1977; Huynh and

Thambirajah, 1984) functions, which are commonly used to model extreme rainfall (Chow et al, 1988). To do this, we will use the data available from the Almodóvar reservoir weather station in the province of Cádiz.

The aim of the present work is to carry out a comparative analysis of these 2 distribution functions and select that which best model the data provided by the weather station, taking into account (i) goodness-of-fit tests, (ii) the cumulative probability and return period obtained to establish design flow discharge in hydro-engineering with a conservative approach (Ministerio

Para la Transición Ecológica y el Reto Demográfico), and (iii) best-performance functions for return periods not exceeding 50 years suitable for estimating mean annual sediment yields (Rodríguez et al, 2022). The goodness-of-fit tests used are the Kolmogorov–Smirnov test (Maity, 2018) and the chi-square test (Huang et al, 2008; Wuensch, 2011). The return periods considered range up to 500 years, which is suitable for most engineering applications (International Commission of Large Dams; Watt and Marsalek, 2013). Furthermore, return periods used to estimate mean interannual sediment yields are up to 50 years, these being the most relevant periods (Mailhot and Duchesne, 2010; Montes-Pajuelo et al, 2024; Rodríguez et al, 2023).

## II. METHODS

In this section we will discuss the weather station from which the study data set was obtained and the distribution functions studied.

### A. Weather station and rainfall dataset

Data are available from Almodovar reservoir weather station located in Cádiz (Andalusia, southern Spain). The annual 24 h maximum rainfall data were obtained from this station, for which data are available from 1990 to 2022. Table 1 indicates the geographical locations of this weather station.

TABLE I  
GEOGRAPHICAL COORDINATES OF THE WEATHER STATION.

Latitude	Longitude	Elevation (m asl)	Interval (years)	Number of data
36° 09' 15"	-5° 39' 03"	105	1990-2022	33

Table 2 presents a statistical summary of the annual maximum 24 h rainfall data set obtained from Almodovar reservoir weather station.

TABLE II  
STATISTICAL VARIABLES OF THE ANNUAL MAXIMUM 24 H RAINFALL DATA SET

Annual mean rainfall (mm)	Standard Deviation (mm)	Coefficient of Variation	Coefficient of Skewness
76.30	30.55	0.40	0.68

### B. Rainfall probability distribution functions

This section present a summary of the formulation of each distribution function. Only accumulated frequency distribution functions and their parameters are explained. In order to fit distributions, parameters were estimated using one the method of moments. Log-Gumbel distribution requires numerical methods for adjustment with regional parameters. To clarify the equations, the parameters for each function are differentiated with an individual index for each case.

#### 1) Log-Gumbel distribution.

The log-Gumbel distribution is a continuous probability distribution of a random variable, defined such that its logarithmic transformation follows a Gumbel distribution; that is, if  $Y = \ln(X)$  follows a Gumbel distribution, then  $X$  follows a log-Gumbel distribution. It is a particular case of the Fréchet

function (extreme value distribution type II) with a position parameter equal to zero (Heo and Salas, 1996). The log-Gumbel distribution is commonly used in rainfall analysis, as precipitation data tend to fit this distribution better after a logarithmic transformation. This is particularly useful for analysing extreme rainfall events, such as high or very low rainfall.

The log-Gumbel cumulative distribution function is:

$$F(x) = e^{-e^{-\frac{-(\ln(x)-\beta)}{\alpha}}}}, \quad x > 0 \quad (1)$$

where  $\alpha > 0$  and  $\beta$  are the scale and location parameters, respectively. To fit the curve, using the method of moments, we obtain the following equations:

$$\bar{x} = \frac{S_x \Gamma(1 - \alpha)}{\sqrt{\Gamma(1 - 2\alpha) - \Gamma^2(1 - \alpha)}} \quad (2)$$

$$\beta = \ln\left(\frac{\bar{x}}{\Gamma(1 - \alpha)}\right) \quad (3)$$

where  $x$  and  $S_x$  are the sample mean and standard deviation, respectively, and  $\Gamma$  is the gamma function. Equation (2) can be solved using a numerical method (e.g., the Newton–Raphson method) to obtain an estimate of the scale parameter  $\alpha$ . An estimate of the location parameter  $\beta$  is then obtained using Equation (3).

#### 2) Log-Pearson III distribution.

The log-Pearson type III distribution is a continuous probability distribution of a random variable whose log transformation follows a Pearson type III distribution; that is, if  $Y = \ln(X)$  follows a Pearson type III distribution, then  $X$  follows a log-Pearson type III distribution (Pearson, 1984). It is used in hydrology for frequency analysis. If the observations present a very high positive skew, then the log-Pearson type III distribution is suitable for modelling. This logarithmic transformation reduces the skewness and can even transform positively skewed data into negatively skewed data (Maity, 2018).

Its cumulative distribution function is given by:

$$F(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_{e^{x_0}}^x \frac{(\ln(t) - x_0)^{\alpha-1} e^{-(\ln(t)-x_0)/\beta}}{t} dt, \quad x \geq e^{x_0} \quad (4)$$

where  $\alpha > 0$ ,  $\beta > 0$ , and  $x_0$  are the shape, rate, and threshold parameters, respectively, and  $\Gamma$  is the gamma function.

The method of moments is used to fit the distribution to the curve, thus obtaining the parametric estimate. If  $z = \ln(x)$  is the logarithmic transformation of the sample, then

$$\alpha = \frac{4}{C_s^2(z)} \quad (5)$$

$$\beta = \frac{C_s(z) S_z}{4} \quad (6)$$

$$x_0 = \bar{z} - \frac{2S_z}{C_s(z)} \quad (7)$$

where  $\bar{z}$ ,  $S_z$ , and  $C_s(z)$  are the mean, standard deviation, and skewness coefficient of the log transformation of the sample, respectively.

III. RESULTS

In this section we will discuss parameter estimation, goodness-of-fit tests and cumulative distribution function frequency analysis. All calculations (parameters calculation, goodness-of-fit tests, graphs, etc.) have been performed using R software (version 4.3.2.) (R Core Team).

A. Parameters calculation and goodness-of-fit tests

Applying equations (1)-(7) to the recorded annual maximum 24-hour rainfall series for the period 1990-2022, the equations for the case studied were obtained. The parameters of each probability distribution function are given in Table 3.

TABLE III  
ADJUSTED PARAMETER VALUES.

<b>Log-Gumbel</b>	$\alpha = 0.23983$	$\beta = 4.14231$	
<b>Log-Pearson III</b>	$\alpha = 26784.68$	$\beta = 0.00244$	$x_0 = -61.16749$

Kolmogorov–Smirnov and chi-square tests were conducted to evaluate the goodness-of-fit of the 2 selected probability distribution functions with respect to the calculated plotting positions through the Weibull formula:  $F(x_i) = \frac{i}{n+1}$ ,  $i = 1, \dots, n$  (where  $x_i$  are the data occupying position  $i$  after incremental sorting of data, and  $n$  is the total number of data). From these goodness-of-fit tests, we obtained the  $p$ -value, that is, the smallest value of the significance level ( $\alpha$ ) at which the hypothesis that the distributions fit the data is rejected. In this study, the significance level was set to 0.05; thus, the hypothesis was accepted if the  $p$ -value was greater than 0.05 and rejected otherwise. As can be seen in Table 4, both distribution functions fit the sample, but the log-Pearson 3 distribution function performs better in both tests.

TABLE IV  
KOLMOGOROV-SMIRNOV AND CHI-SQUARE TEST P-VALUES

	<b>Log-Gumbel</b>	<b>Log-Pearson III</b>
<b>Kolmogorov-Smirnov <math>p</math>-value</b>	0.3789	0.9946
<b>Chi-square <math>p</math>-value</b>	0.0977	0.4706

In Figure 1, we can see graphically the fit of the curves to the data where the Weibull formula has been used for the empirical distribution.

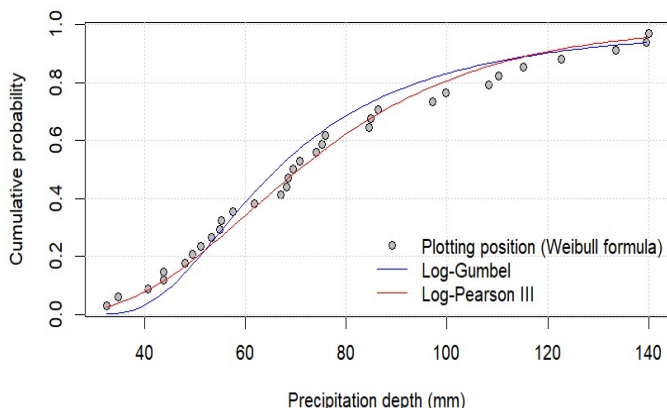


Fig. 1. Cumulative probability versus precipitation depth for the probability distribution functions.

B. Cumulative distribution function frequency analysis

The return period ( $T$ ) can be defined as the average time lapse for an event of a given magnitude to be equalled or exceeded in a statistical sense, for example, when the average period that must elapse before the maximum annual value of the daily precipitation depth at a given geographical point is equalled or exceeded.

Let  $X$  be the random variable representing the annual maximum 24 h rainfall and  $F$  be its cumulative distribution function. Suppose a certain rainfall event whose precipitation depth value is  $x$  mm occurs. Then,  $1 - F(x)$  is the probability that the event will be exceeded in a year. If we now consider the random variable representing the number of years until this event is exceeded ( $\tau$ ), it is easy to see that  $\tau$  follows a geometric distribution with parameter  $1 - F(x)$ . Thus, the return period ( $T$ ) for this rainfall event will be the mean of  $\tau$ ; that is,

$$T = \bar{\tau} = \frac{1}{1 - F(x)} \tag{8}$$

From Equation (8), we obtain

$$x = F^{-1}\left(1 - \frac{1}{T}\right) \tag{9}$$

which represents the precipitation depth as a function of the return period.

Equation (9) was used for the cumulative distribution functions examined. Figure 2 show the application of this equation for  $T$  from 1.2 to 500 years.

In hydraulic, hydrological and sedimentological engineering, one of the most well-established criteria for decision-making is based on adopting those predictions that are on the safe or conservative side. For example, choose the prediction of those functions that provide the highest values of precipitation depth given a return period value.

For some type of hydraulic infrastructure, the distribution that gives a greater precipitation depth for high return period values ( $T > 100$  years) is preferred. In this case, it can be seen in the previous figure that the log-Gumbel function performs better, but it can also be seen that it has high sensitivity. For this reason, its use is discouraged as, due to its sensitivity, it can produce results with large variations when new data are added, thus distorting them in one direction or another.

In the case of calculating the mean interannual sediment

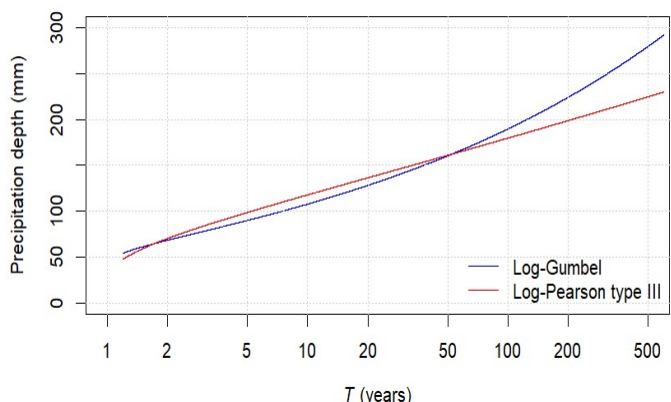


Fig. 2. Precipitation depth against return period for the probability distribution functions and T from 1.2 to 500 years distribution functions.

yield, the function that gives a greater precipitation depth for low return period values ( $T < 50$  years) is preferred. In this case, it can be seen in the previous figure that the log-Pearson function is the one that gives values on the conservative side.

#### IV. CONCLUSIONS

In this work, a comparative analysis of the Log-Gumbel and Log-Pearson III distribution functions has been carried out, with the following conclusions:

Both functions had performed well in the goodness-of-fit tests (with log-Pearson III distribution noticeably better).

For high return period values ( $T > 100$  years), the log-Gumbel function performs better, but it has a high sensitivity, so its use is not recommended (if sensitivity is of interest).

For low return period values ( $T < 50$  years), the log-Pearson III function is preferred.

#### REFERENCES

- Bobee, B.B., Robitaille, R. The use of the Pearson type 3 and log Pearson type 3 distributions revisited. *Water Resour. Res.* 1977, 13,427–443. <https://doi.org/10.1029/WR013i002p00427>
- Chow, V.T., Maidment, D.R., Mays, L.W. *Applied Hydrology*, McGraw-Hill: New York, NY, USA, 1988.
- Coronado-Hernández, Ó. E., Merlano-Sabalza, E., Díaz-Vergara, Z., Coronado-Hernández, J. R. Selection of hydrological probability distributions for extreme rainfall events in the regions of Colombia. *Water*, 12(5), 2020, 1397. <https://doi.org/10.3390/w12051397>
- Gumbel, E.J. Les valeurs extrêmes des distributions statistiques. *Ann. De L'institut Henri Poincaré* 1935, 5, 115–158. Available online: [http://www.numdam.org/item/AIHP\\_1935\\_5\\_2\\_115\\_0.pdf](http://www.numdam.org/item/AIHP_1935_5_2_115_0.pdf)
- Gumbel, E.J. The return period of flood flows. *Ann. Math. Stat.* 1941, 12, 163–190. Available online: <https://www.jstor.org/stable/2235766>
- Heo, J.H., Salas, J.D. Estimation of quantiles and confidence intervals for the log-Gumbel distribution. *Stoch. Hydrol. Hydraul.* 1996, 10, 187–207. <https://doi.org/10.1007/BF01581463>
- Hossain, F., Jeyachandran, I., Pielke, R., Sr. Dam safety effects due to human alteration of extreme precipitation. *Water Resour. Res.*, 2010, 46(3). <https://doi.org/10.1029/2009WR007704>
- Huang, Y. P., Lee, C. H., Ting, C. S. Improved estimation of hydrologic data using the chi-square goodness-of-fit test. *J. Chin. Inst. Eng.*, 31(3), 2008, 515-521. <https://doi.org/10.1080/02533839.2008.9671406>
- Huynh, N.P., Thambirajah, J.A. Applications of the log Pearson type-3 distribution in hydrology. *J. Hydrol.* 1984, 73, 359–372.
- International Commission of Large Dams. Definition of a Large Dam. Available online: [https://www.icold-cigb.org/GB/dams/definition\\_of\\_a\\_large\\_dam.asp](https://www.icold-cigb.org/GB/dams/definition_of_a_large_dam.asp)
- Mailhot, A., Duchesne, S. Design criteria of urban drainage infrastructures under climate change. *J. Water Resour. Plan. Manag.* 2010, 136, 201–208. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000023](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000023)
- Maity, R. *Statistical Methods in Hydrology and Hydroclimatology*; Springer: Singapore, 2018; Volume 555. <https://doi.org/10.1007/978-981-16-5517-3>
- Ministerio Para la Transición Ecológica y el Reto Demográfico. Gobierno de España. Agencia Estatal de Meteorología (AEMET). Available online: <https://www.aemet.es/en/serviciosclimaticos>
- Montes-Pajuelo, R., Rodríguez-Pérez, Á. M., López, R., Rodríguez, C. A. *Analysis of Probability Distributions for Modelling Extreme Rainfall Events and Detecting Climate Change: Insights from Mathematical and Statistical Methods*. *Mathematics*, 2024, 12(7), 1093. <https://doi.org/10.3390/math12071093>
- Pearson, K. Contributions to the mathematical theory of evolution. *Philos. Trans. R. Soc.* 1894, 185, 71–110. <https://doi.org/10.1098/rsta.1895.0010>
- R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023. Available online: <https://www.R-project.org>
- Rodríguez, C.A., Rodríguez-Pérez, Á.M., Mancera, J. J. C., Torres, J. A. H., Carmona, N. G., & Bahamonde, M. I. Applied methodology based on HEC-HMS for reservoir filling estimation due to soil erosion. *Journal of Hydrology and Hydromechanics*, 2022, 70(3), 341-356. <https://doi.org/10.2478/johh-2022-0020>
- Rodríguez, C.A., Rodríguez-Pérez, Á.M., López, R., Hernández-Torres, J.A., Caparrós-Mancera, J.J. Sensitivity Analysis in Mean Annual Sediment Yield Modeling with Respect to Rainfall Probability Distribution Functions. *Land*, 2023, 12, 35. <https://doi.org/10.3390/land12010035>
- Teng, F., Huang, W., Ginis, I. Hydrological modeling of storm runoff and snowmelt in Taunton River Basin by applications of HEC-HMS and PRMS models. *Natural Hazards*, 2018, 91(1), 179-199. <https://doi.org/10.1007/s11069-017-3121-y>
- Watt, E., Marsalek, J. Critical review of the evolution of the design storm event concept. *Can. J. Civ. Eng.* 2013, 40, 105–113. <https://doi.org/10.1139/cjce-2011-0594>
- Wuensch, K.L. *Chi-Square Tests*. In *International Encyclopedia of Statistical Science*; Lovric, M., Ed.; Springer: Berlin/Heidelberg, Germany, 2011. [https://doi.org/10.1007/978-3-642-04898-2\\_173](https://doi.org/10.1007/978-3-642-04898-2_173)



**Reconocimiento – NoComercial (by-nc):** Se permite la generación de obras derivadas siempre que no se haga un uso comercial. Tampoco se puede utilizar la obra original con finalidades comerciales.